

**APPLICATION
FOR
UNITED STATES LETTERS PATENT
BY
JORDAN J.N. TANG
LIN HONG
AND
ARUN K. GHOSH
FOR
INHIBITORS OF MEMAPSIN 2 AND USE THEREOF**

INHIBITORS OF MEMAPSIN 2 AND USE THEREOF

Background of the Invention

This application claims priority to U.S.S.N. 60/141,363 filed June 28, 1999 by Lin, et al., U.S.S.N. 60/168,060 filed November 30, 1999 by Lin, et al., 5 U.S.S.N. 60/177,836 filed January 25, 2000 by Lin, et al., U.S.S.N. 60/178,368 filed January 27, 2000 by Lin, et al., and U.S.S.N. 60/210,292 filed June 8, 2000 by Lin Hong, et al., the teachings of which are incorporated by reference herein.

10 This invention is in the area of the design and synthesis of specific inhibitors of the aspartic protease Memapsin 2 (beta-secretase) which are useful in the treatment and/or prevention of Alzheimer's Disease.

15 Alzheimer's disease (AD) is a degenerative disorder of the brain first described by Alios Alzheimer in 1907 after examining one of his patients who suffered drastic reduction in cognitive abilities and had generalized dementia (*The early story of Alzheimer's Disease*, edited by Bick *et al.* (Raven Press, New York 1987)). It is the leading cause of dementia in elderly persons. AD patients have increased problems with memory loss and intellectual functions which progress to the point where they cannot function as normal individuals. With the loss of intellectual skills the patients exhibit personality changes, socially inappropriate actions and schizophrenia (*A Guide to the Understanding of 20 Alzheimer's Disease and Related Disorders*, edited by Jorm (New York University Press, New York 1987)). AD is devastating for both victims and their families, for there is no effective palliative or preventive treatment for the inevitable neurodegeneration.

25 AD is associated with neuritic plaques measuring up to 200 μm in diameter in the cortex, hippocampus, subiculum, hippocampal gyrus, and amygdala. One of the principal constituents of neuritic plaques is amyloid, which is stained by Congo Red (Fisher (1983); Kelly Microbiol. Sci. 1(9):214-219 (1984)). Amyloid plaques stained by Congo Red are extracellular, pink or rust-colored in bright field, and birefringent in polarized light. The plaques are 30 composed of polypeptide fibrils and are often present around blood vessels, reducing blood supply to various neurons in the brain.

Various factors such as genetic predisposition, infectious agents, toxins, metals, and head trauma have all been suggested as possible mechanisms of AD neuropathy. Available evidence strongly indicates that there are distinct types of genetic predispositions for AD. First, molecular analysis has provided evidence for mutations in the amyloid precursor protein (APP) gene in certain AD-stricken families (Goate *et al.* *Nature* 349:704-706 (1991); Murrell *et al.* *Science* 254:97-99 (1991); Chartier-Harlin *et al.* *Nature* 353:844-846 (1991); Mullan *et al.*, *Nature Genet.* 1:345-347 (1992)). Additional genes for dominant forms of early onset AD reside on chromosome 14 and chromosome 1 (Rogaev *et al.*, *Nature* 376:775-778 (1995); Levy-Lahad *et al.*, *Science* 269:973-977 (1995); Sherrington *et al.*, *Nature* 375:754-760 (1995)). Another loci associated with AD resides on chromosome 19 and encodes a variant form of apolipoprotein E (Corder, *Science* 261:921-923 (1993)).

Amyloid plaques are abundantly present in AD patients and in Down's Syndrome individuals surviving to the age of 40. The overexpression of APP in Down's Syndrome is recognized as a possible cause of the development of AD in Down's patients over thirty years of age (Rumble *et al.*, *New England J. Med.* 320:1446-1452 (1989); Mann *et al.*, *Neurobiol. Aging* 10:397-399 (1989)). The plaques are also present in the normal aging brain, although at a lower number. These plaques are made up primarily of the amyloid β peptide (A β ; sometimes also referred to in the literature as β -amyloid peptide or β peptide) (Glenner and Wong, *Biochem. Biophys. Res. Comm.* 120:885-890 (1984)), which is also the primary protein constituent in cerebrovascular amyloid deposits. The amyloid is a filamentous material that is arranged in beta-pleated sheets. A β is a hydrophobic peptide comprising up to 43 amino acids.

The determination of its amino acid sequence led to the cloning of the APP cDNA (Kang *et al.*, *Nature* 325:733-735 (1987); Goldgaber *et al.*, *Science* 235:877-880 (1987); Robakis *et al.*, *Proc. Natl. Acad. Sci.* 84:4190-4194 (1987); Tanzi *et al.*, *Nature* 331:528-530 (1988)) and genomic APP DNA (Lemaire *et al.*, *Nucl. Acids Res.* 17:517-522 (1989); Yoshikai *et al.*, *Gene* 87, 257-263 (1990)). A number of forms of APP cDNA have been identified, including the

three most abundant forms, APP695, APP751, and APP770. These forms arise from a single precursor RNA by alternate splicing. The gene spans more than 175 kb with 18 exons (Yoshikai *et al.* (1990)). APP contains an extracellular domain, a transmembrane region and a cytoplasmic domain. A β consists of up to 28 amino acids just outside the hydrophobic transmembrane domain and up to 15 residues of this transmembrane domain. A β is normally found in brain and other tissues such as heart, kidney and spleen. However, A β deposits are usually found in abundance only in the brain.

Van Broeckhoven *et al.*, *Science* 248:1120-1122 (1990), have demonstrated that the APP gene is tightly linked to hereditary cerebral hemorrhage with amyloidosis (HCHWA-D) in two Dutch families. This was confirmed by the finding of a point mutation in the APP coding region in two Dutch patients (Levy *et al.*, *Science* 248:1124-1128 (1990)). The mutation substituted a glutamine for glutamic acid at position 22 of the A β (position 618 of APP695, or position 693 of APP770). In addition, certain families are genetically predisposed to Alzheimer's disease, a condition referred to as familial Alzheimer's disease (FAD), through mutations resulting in an amino acid replacement at position 717 of the full length protein (Goate *et al.* (1991); Murrell *et al.* (1991); Chartier-Harlin *et al.* (1991)). These mutations co-segregate with the disease within the families and are absent in families with late-onset AD. This mutation at amino acid 717 increases the production of the A β_{1-42} form of A β from APP (Suzuki *et al.*, *Science* 264:1336-1340 (1994)). Another mutant form contains a change in amino acids at positions 670 and 671 of the full length protein (Mullan *et al.* (1992)). This mutation to amino acids 670 and 671 increases the production of total A β from APP (Citron *et al.*, *Nature* 360:622-674 (1992)).

APP is processed *in vivo* at three sites. The evidence suggests that cleavage at the β -secretase site by a membrane associated metalloprotease is a physiological event. This site is located in APP 12 residues away from the luminal surface of the plasma membrane. Cleavage of the β -secretase site (28 residues from the plasma membrane's luminal surface) and the β -secretase site

(in the transmembrane region) results in the 40/42-residue β -amyloid peptide (A β), whose elevated production and accumulation in the brain are the central events in the pathogenesis of Alzheimer's disease (for review, see Selkoe, D.J. *Nature* 399:23-31 (1999)). Presenilin 1, another membrane protein found in 5 human brain, controls the hydrolysis at the APP (β -secretase site and has been postulated to be itself the responsible protease (Wolfe, M.S. et al., *Nature* 398:513-517 (1999)). Presenilin 1 is expressed as a single chain molecule and its processing by a protease, presenilinase, is required to prevent it from rapid degradation (Thinakaran, G. et al., *Neuron* 17:181-190 (1996) and Podlisny, 10 M.B., et al., *Neurobiol. Dis.* 3:325-37 (1997)). The identity of presenilinase is unknown. The *in vivo* processing of the β -secretase site is thought to be the rate-limiting step in A β production (Sinha, S. & Lieberburg, I., *Proc. Natl. Acad. Sci., USA*, 96:11049-11053 (1999)), and is therefore a strong therapeutic target.

15 The design of inhibitors effective in decreasing amyloid plaque formation is dependent on the identification of the critical enzyme(s) in the cleavage of APP to yield the 42 amino acid peptide, the A β_{1-42} form of A β . Although several enzymes have been identified, it has not been possible to produce active enzyme. Without active enzyme, one cannot confirm the 20 substrate specificity, determine the subsite specificity, nor determine the kinetics or critical active site residues, all of which are essential for the design of inhibitors.

25 Memapsin 2 has been shown to be beta-secretase, a key protease involved in the production in human brain of beta-amyloid peptide from beta-amyloid precursor protein (for review, see Selkoe, D.J. *Nature* 399:23-31 (1999)). It is now generally accepted that the accumulation of beta-amyloid peptide in human brain is a major cause for the Alzheimer's disease. Inhibitors specifically designed for human memapsin 2 should inhibit or decrease the 30 formation of beta-amyloid peptide and the progression of the Alzheimer's disease.

Memapsin 2 belongs to the aspartic protease family. It is homologous in amino acid sequence to other eukaryotic aspartic proteases and contains motifs specific to that family. These structural similarities predict that memapsin 2 and other eukaryotic aspartic proteases share common catalytic mechanism Davies, D.R., *Annu. Rev. Biophys. Chem.* 19, 189 (1990). The most successful inhibitors for aspartic proteases are mimics of the transition state of these enzymes. These inhibitors have substrate-like structure with the cleaved planar peptide bond between the carbonyl carbon and the amide nitrogen replaced by two tetrahedral atoms, such as hydroxyethylene [-CH(OH)-CH₂-], which was originally discovered in the structure of pepstatin (Marciniszyn et al., 1976).

However, for clinical use, it is preferable to have small molecule inhibitors which will pass through the blood brain barrier and which can be readily synthesized. It is also desirable that the inhibitors are relatively inexpensive to manufacture and that they can be administered orally. Screening of thousands of compounds for these properties would require an enormous effort. To rationally design memapsin 2 inhibitors for treating Alzheimer's disease, it will be important to know the three-dimensional structure of memapsin 2, especially the binding mode of an inhibitor in the active site of this protease.

It is therefore an object of the present invention to provide purified, recombinant, and active memapsin 2, as well as its substrate and subsite specificity and critical active site residues.

It is a further object of the present invention to provide compositions and methods for synthesis of inhibitors of memapsin 2.

It is a still further object of the present invention to provide compositions that interact with memapsin 2 or its substrate to inhibit cleavage by the memapsin 2 which can cross the blood brain barrier (BBB).

It is therefore an object of the present invention to provide means for rational design and screening of compounds for inhibition of memapsin 2.

Summary of the Invention

Methods for the production of purified, catalytically active, recombinant memapsin 2 have been developed. The substrate and subsite specificity of the catalytically active enzyme have been determined. The active enzyme and 5 assays for catalytic activity are useful in screening libraries for inhibitors of the enzyme.

The substrate and subsite specificity information was used to design substrate analogs of the natural memapsin 2 substrate that can inhibit the function of memapsin 2. The substrate analogs are based on peptide sequences, 10 shown to be related to the natural peptide substrates for memapsin 2. The substrate analogs contain at least one analog of an amide (peptide) bond which is not capable of being cleaved by memapsin 2. Processes for the synthesis of two substrate analogues including isosteres at the sites of the critical amino acid residues were developed and the substrate analogues, OMR99-1 and OM99-2, 15 were synthesized. OM99-2 is based on an octapeptide Glu-Val-Asn-Leu-Ala-Ala-Glu-Phe (SEQ ID NO:28) with the Leu-Ala peptide bond substituted by a transition-state isostere hydroxyethylene group. The inhibition constant of OM99-2 is 1.6×10^{-9} M against recombinant pro-memapsin 2. Crystallography of memapsin 2 bound to this inhibitor was used to determine the three 20 dimensional structure of the protein, as well as the importance of the various residues in binding.

This information can be used by those skilled in the art to design new inhibitors, using commercially available software programs and techniques familiar to those in organic chemistry and enzymology, to design new 25 inhibitors. For example, the side chains of the inhibitors may be modified to produce stronger interactions (through hydrogen bonding, hydrophobic interaction, charge interaction and/or van der Waal interaction) in order to increase inhibition potency. Based on this type of information, the residues with minor interactions may be eliminated from the new inhibitor design to decrease 30 the molecular weight of the inhibitor. The side chains with no structural hindrance from the enzyme may be cross-linked to lock in the effective

inhibitor conformation. This type of structure also enables the design of peptide surrogates which may effectively fill the binding sites of memapsin 2 yet produce better pharmaceutical properties.

The examples demonstrate the production of catalytically active enzyme,
5 design and synthesis of inhibitors, and how the crystal structure was obtained. The examples thereby demonstrate how the methods and materials described herein can be used to screen libraries of compounds for other inhibitors, as well as for design of inhibitors. These inhibitors are useful in the prevention and/or treatment of Alzheimer's disease as mediated by the action of the beta secretase
10 memapsin 2, in cleaving APP.

Brief Description of the Drawings

Figure 1 depicts the plasmid construct of vector pET-11a-memapsin 2-T1 and pET-11a-memapsin 2-T2. The T7 promotor, amino acid sequence from the vector (T7 protein) (SEQ ID NO:3), and the beginning and ending of the
15 memapsin 2 T1 and T2 construct are shown. Construct promemapsin 2-T1 was used in the preparation of protein for crystallization and includes residues 1v-15v which are derived from vector pET-11a. Residues 1p-48p are putative pro-peptide. Residues 1-393 correspond to the mature protease domain and C-terminal extension. The residue numbering of memapsin 2 starts at the aligned
20 N-terminal position of pepsin (Figure 3).

Figure 2A is a graph of the initial rate of hydrolysis of synthetic peptide swAPP (see Table 1) by M2_{pd} at different pH. Figure 2B is a graph of the relative k_{cat}/K_m values for steady-state kinetic of hydrolysis of peptide substrates by M2_{pd}.

25 Figures 3A and 3B are the chemical structures of memapsin 2 inhibitors, OM99-2 and OM99-1.

Figure 4A is a graph of the inhibition of recombinant memapsin 2 by OM99-1. Figure 4B is a graph of the inhibition of recombinant memapsin 2 by OM99-2.

30 Figures 5A-E are photographs of crystals of recombinant memapsin 2-OM99-2 complex.

Figure 6 is a stereo view of crystal structure of memapsin 2 protease domain with bound OM99-2. The polypeptide backbone of memapsin 2 is shown as a ribbon diagram. The N-lobe and C-lobe are blue and yellow, respectively, except the insertion loops (designated A to G, see Figure 2) in the C-lobe are magenta and the C-terminal extension is green. The inhibitor bound between the lobes is shown in red.

Figure 7 is a stereo view of comparison of the three-dimensional structures of memapsin 2 and pepsin. The molecular surface of the former is significantly larger by the insertion of surface loops and helix and the C-terminal extension. Chain tracing of human memapsin 2 is dark blue and is grey for human pepsin. The light blue balls represent identical residues which are topologically equivalent. The disulfide bonds are shown in red for memapsin 2 and orange for pepsin. The C-terminal extension is in green.

Figure 8 is a schematic presentation of interaction between OM99-2 and memapsin 2 protease domain. The S₃' and S₄' subsites are not defined.

Figure 9 is a stereo presentation of interactions between inhibitor OM99-2 (orange) and memapsin 2 (light blue). Nitrogen and oxygen atoms are marked blue and red, respectively. Hydrogen bonds are indicated in yellow dotted lines. Memapsin 2 residues which comprise the binding subsites are included. Residues P₄, P₃, P₂, P₁ and P₁' (defined in Figure 8) of OM99-2 are in an extended conformation. Inhibitor chain turns at residue P2' which makes a distinct kink at this position. The backbone of P₃' and P₄' directs the inhibitor to exit the active site.

Figure 10 are schematics of the cross linking between P₃ Val and P₁ Leu side chains in the design of new inhibitors for memapsin 2 based on the current crystal structure. R and R' at positions P₂ and P₁' indicate amino acid side chains. Other structural elements of inhibitor are omitted for clarity.

Figure 11 are schematics of the cross linking between P₄ Glu and P₂ Asn side chains in the design of new inhibitors for memapsin 2 based on the current crystal structure. R at position P₃ indicates amino acid side chain. Other structural elements of inhibitor are omitted for clarity.

Figure 12 is a schematic of the design for the side chain at the P_{1'} subsite for the new memapsin 2 inhibitors based on the current crystal structure. Arrows indicate possible interactions between memapsin 2 and inhibitor. Other structural elements of inhibitor are omitted for clarity.

5 Figure 13 is a schematic of the design of two six-membered rings in the inhibitor structure by the addition of atoms A and B. The ring formation involves the P_1 -Leu side chain the the peptide backbone near P_1 , P_2 , and P_3 . The new bonds are in dotted lines. A methyl group can be added to the beta-carbon of P_1 -Leu. Other structural elements of inhibitor are omitted for clarity.

10 **Detailed Description of the Invention**

I. Preparation of Catalytically Active Recombinant Memapsin 2

Cloning and Expression of Memapsin 2

Memapsin 2 was cloned and the nucleotide (SEQ ID NO. 1) and predicted amino acid (SEQ ID NO. 2) sequences were determined, as described in Example 1. The cDNA was assembled from the fragments. The nucleotide and the deduced protein sequence are shown in SEQ ID NOs. 1 and 2, respectively. The protein is the same as the aspartic proteinase 2 (ASP2) described in EP 0 855 444 A by SmithKline Beecham Pharmaceuticals, (published July 29, 1998), and later reported by Sinha, et al., Nature 402, 537-540 (December 1999) and Vassar, et al., Science 286, 735-741 (22 October 1999).

Pro-memapsin 2 is homologous to other human aspartic proteases. Based on the alignments, Pro-memapsin 2 contains a *pro* region, an aspartic protease region, and a trans-membrane region near the C-terminus. The C-terminal domain is over 80 residues long. The active enzyme is memapsin 2 and its pro-enzyme is pro-memapsin 2.

Refolding Catalytically Active Enzyme

In order to determine the substrate specificity and to design inhibitors, it is necessary to express catalytically active recombinant enzyme. No other known proteases contain a transmembrane domain. The presence of transmembrane domains makes the recombinant expression of these proteins

less predictable and more difficult. The transmembrane region often needs to be removed so that secretion of the protein can take place. However, the removal of the transmembrane region can often alter the structure and/or function of the protein.

5 The starting assumption was that the region of memapsin 2 that is homologous with other aspartic proteases would independently fold in the absence of the transmembrane domain, and would retain protease activity in the absence of the C-terminal transmembrane region. The transmembrane region appears to serve as a membrane anchor. Since the active site is not in the

10 transmembrane region and activity does not require membrane anchoring, memapsin 2 was expressed in *E. coli* in two different lengths, both without the transmembrane region, and purified, as described in Example 3. The procedures for the culture of transfected bacteria, induction of synthesis of recombinant proteins and the recovery and washing of inclusion bodies containing

15 recombinant proteins are essentially as described by Lin et al., (1994). Refolding was not a simple matter, however. Two different refolding methods both produced satisfactory results. In both methods, the protein was dissolved in a strong denaturing/reducing solution such as 8 M urea/100 mM beta-mercaptoethanol. The rate at which the protein was refolded, and in what

20 solution, was critical to activity. In one method, the protein is dissolved into 8 M urea/100 mM beta-mercaptoethanol then rapidly diluted into 20 volumes of 20 mM-Tris, pH 9.0, which is then slowly adjusted to pH 8 with 1 M HCl. The refolding solution was then kept at 4° C for 24 to 48 hours before proceeding with purification. In the second method, an equal volume of 20 mM Tris, 0.5

25 mM oxidized/1.25 mM reduced glutathione, pH 9.0 is added to rapidly stirred pro-memapsin 2 in 8 M urea/10 mM beta-mercaptoethanol. The process is repeated three more times with 1 hour intervals. The resulting solution is then dialyzed against sufficient volume of 20 mM Tris base so that the final urea concentration is 0.4 M. The pH of the solution is then slowly adjusted to 8.0

30 with 1 M HCl.

The refolded protein is then further purified by column chromatography, based on molecular weight exclusion, and/or elution using a salt gradient, and analyzed by SDS-PAGE analysis under reduced and non-reduced conditions.

II. Substrate Specificity and Enzyme Kinetics of Memapsin 2

5 Substrate Specificity

The tissue distribution of the memapsin 2 was determined, as described in Example 2. The presence of memapsin 2 (M2) in the brain indicated that it might hydrolyze the β -amyloid precursor protein (APP). As described below, detailed enzymatic and cellular studies demonstrated that M2 fits all the criteria

10 of the β -secretase.

The M2 three-dimensional structure modeled as a type I integral membrane protein. The model suggested that its globular protease unit can hydrolyze a membrane anchored polypeptide at a distance range of 20-30 residues from the membrane surface. As a transmembrane protein of the brain, 15 APP is a potential substrate and its beta-secretase site, located about 28 residues from the plasma membrane surface, is within in the range for M2 proteolysis.

A synthetic peptide derived from this site (SEVKM/DAEFR) (SEQ ID NO:4) was hydrolyzed by M2_{pd} (modified M2 containing amino acids from Ala^{8P} to Ala³²⁶) at the beta-secretase site (marked by /). A second peptide 20 (SEVNL/DAEFR) (SEQ ID NO:5) derived from the APP beta-secretase site and containing the 'Swedish mutation' (Mullan, M. *et al.*, *Nature Genet.* 2:340-342 (1992)), known to elevate the level of alpha-beta production in cells (Citron, M. *et al.*, *Nature* 260:672-674 (1992)), was hydrolyzed by M2_{pd} with much higher catalytic efficiency. Both substrates were optimally cleaved at pH 4.0. A 25 peptide derived from the processing site of presenilin 1 (SVNM/AEGD) (SEQ ID NO:6) was also cleaved by M2_{pd} with less efficient kinetic parameters. A peptide derived from the APP gamma-secretase site (KGGVVIATVIVK) (SEQ ID NO:7) was not cleaved by M2_{pd}. Pepstatin A inhibited M2_{pd} poorly (IC₅₀ approximately approximately 0.3 mM). The kinetic parameters indicate that 30 both presenilin 1 (k_{cat} , 0.67 s⁻¹; K_m, 15.2 mM; k_{cat}/K_m , 43.8 s⁻¹M⁻¹) and native

APP peptides (k_{cat}/K_m , $39.9\text{ s}^{-1}\text{M}^{-1}$) are not as good substrates as the Swedish APP peptide (k_{cat} , $2.45\text{ s}^{-1}, K_m$, 1 mM ; k_{cat}/K_m , $2450\text{ s}^{-1}\text{M}^{-1}$).

To determine if M2 possesses an APP beta-secretase function in mammalian cells, memapsin 2 was transiently expressed in HeLa cells (Lin, X., et al., *FASEB J.* 7:1070-1080 (1993)), metabolically pulse-labeled with ^{35}S -Met, then immunoprecipitated with anti-APP antibodies for visualization of APP-generated fragments after SDS-polyacrylamide electrophoresis and imaging. SDS-PAGE patterns of immuno-precipitated APP N β -fragment (97 kD band) from the conditioned media (2 h) of pulse-chase experiments showed that APP was cleaved by M2. Controls transfected with APP alone and co-transfected with APP and M2 with Bafilomycin A1 added were performed. SDS-PAGE patterns of APP β C-fragment (12 kD) were immunoprecipitated from the conditioned media of the same experiment as discussed above. Controls transfected with APP alone; co-transfected with APP and M2; co-transfected with APP and M2 with Bafilomycin A1; transfections of Swedish APP; and co-transfections of Swedish APP and M2 were performed. SDS-PAGE gels were also run of immuno-precipitated M2 (70 kD), M2 transfected cells; untransfected HeLa cells after long time film exposure; and endogenous M2 from HEK 293 cells. SDS-PAGE patterns of APP fragments (100 kD betaN-fragment and 95 kD betaN-fragment) recovered from conditioned media after immuno-precipitation using antibodies specific for different APP regions indicated that memapsin 2 cleaved APP.

Cells expressing both APP and M2 produced the 97 kD APP beta N-fragment (from the N-terminus to the beta-secretase site) in the conditioned media and the 12 kD betaC-fragment (from the beta-secretase site to the C-terminus) in the cell lystate. Controls transfected with APP alone produced little detectable betaN-fragment and no beta C-fragment. Bafilomycin A1, which is known to raise the intra-vesicle pH of lysosomes/endosomes and has been shown to inhibit APP cleavage by beta-secretase (Knops, J. et al., *J. Biol. Chem.* 270:2419-2422 (1995)), abolished the production of both APP fragments beta N- and beta C- in co-transfected cells. Cells transfected with Swedish APP

alone did not produce the beta C-fragment band in the cell lysate but the co-transfection of Swedish APP and M2 did. This Swedish beta C-fragment band is more intense than that of wild-type APP. A 97-kD beta N-band is also seen in the conditioned media but is about equal intensity as the wild-type APP

5 transfection.

These results indicate that M2 processes the beta-secretase site of APP in acidic compartments such as the endosomes. To establish the expression of transfected M2 gene, the pulse-labeled cells were lysed and immunoprecipitated by anti-M2 antibodies. A 70 kD M2 band was seen in cells 10 transfected with M2 gene, which has the same mobility as the major band from HEK 293 cells known to express beta-secretase (Citron, M. et al., *Nature* 260:672-674 (1992)). A very faint band of M2 is also seen, after a long film exposure, in untransfected HeLa cells, indicating a very low level of endogenous M2, which is insufficient to produce betaN- or betaC-fragments without M2 15 transfection. Antibody alpha-beta₁₋₁₇, which specifically recognizes residues 1-17 in alpha-beta peptide, was used to confirm the correct beta-secretase site cleavage. In cells transfected with APP and M2, both beta N- and beta N- fragments are visible using an antibody recognizing the N-terminal region of APP present in both fragments. Antibody Abeta₁₋₁₇ recognize the beta N- 20 fragment produced by endogenous beta-secretase in the untransfected cells. This antibody was, however, unable to recognize the betaN-fragment known to be present in cells co-transfected with APP and M2. These observations confirmed that betaN-fragment is the product of beta-secretase site cut by M2, which abolished the recognition epitope of alpha-beta₁₋₁₇.

25 The processing of APP by M2 predicts the intracellular colocalization of the two proteins. HeLa cells co-expressing APP and M2 were stained with antibodies directed toward APP and M2 and visualized simultaneously by CSLM using a 100x objective. Areas of colocalization appeared in yellow.

30 Immunodetection observed by confocal microscopy of both APP and M2 revealed their colocalization in the superimposed scans. The distribution of both

proteins is consistent with their residence in lysosomal/endosomal compartments.

In specificity studies, it was found that M2_{pd} cleaved its *pro* peptide (2 sites) and the protease portion (2 sites) during a 16 h incubation after activation 5 (Table 1). Besides the three peptides discussed above, M2_{pd} also cleaved oxidized bovine insulin B chain and a synthetic peptide Nch. Native proteins were not cleaved by M2_{pd}.

The data indicate that human M2 fulfills all the criteria of a beta-secretase which cleaves the beta-amyloid precursor protein (APP): (a) M2 and 10 APP are both membrane proteins present in human brain and co-localize in mammalian cells, (b) M2 specifically cleaves the beta-secretase site of synthetic peptides and of APP in cells, (c) M2 preferentially cleaves the beta-secretase site from the Swedish over the wild-type APP, and (d) the acidic pH optimum for 15 M2 activity and bafilomycin A1 inhibition of APP processing by M2 in the cells are consistent with the previous observations that beta-secretase cleavage occurs in acidic vesicles (Knops, J., et al., *J. Biol. Chem.* 270:2419-2422 (1995)). The spontaneous appearance of activity of recombinant pro-M2 in an acidic solution suggests that, intracellularly, this zymogen can by itself generate activity in an acidic vesicle like an endosome.

20 **II. Design and Synthesis of Inhibitors**

Design of Substrate Analogs for Memapsin 2.

The five human aspartic proteases have homologous amino acid sequences and have similar three-dimensional structures. There are two aspartic residues in the active site and each residue is found within the signature aspartic 25 protease sequence motif, Asp-Thr/Ser-Gly- (SEQ ID NO:8). There are generally two homologous domains within an aspartic protease and the substrate binding site is positioned between these two domains, based on the three-dimensional structures. The substrate binding sites of aspartic proteases generally recognize eight amino acid residues. There are generally four residues 30 on each side of the amide bond which is cleaved by the aspartic protease.

Typically the side chains of each amino acid are involved in the specificity of the substrate/aspartic protease interaction. The side chain of each substrate residue is recognized by regions of the enzyme which are collectively called sub-sites. The generally accepted nomenclature for the protease sub-sites and their corresponding substrate residues are shown below, where the double slash represents the position of bond cleavage.

5 Protease sub-sites S4 S3 S2 S1 S1' S2' S3' S4'
Substrate residues P4 P3 P2 P1 // P1' P2' P3' P4'

While there is a general motif for aspartic protease substrate recognition, 10 each protease has a very different substrate specificity and breadth of specificity. Once the specificity of an aspartic protease is known, inhibitors can be designed based on that specificity, which interact with the aspartic protease in a way that prevents natural substrate from being efficiently cleaved. Some aspartic proteases have specificities which can accommodate many different residues in 15 each of the sub-sites for successful hydrolysis. Pepsin and cathepsin D have this type of specificity and are said to have "broad" substrate specificity. When only a very few residues can be recognized at a sub-site, such as in renin, the aspartic protease is said to have a stringent or narrow specificity.

The information on the specificity of an aspartic protease can be used to 20 design specific inhibitors in which the preferred residues are placed at specific sub-sites and the cleaved peptide bond is replaced by an analog of the transition-state. These analogs are called transition state isosteres. Aspartic proteases cleave amide bonds by a hydrolytic mechanism. This reaction mechanism involves the attack by a hydroxide ion on the β -carbon of the amino acid.

25 Protonation must occur at the other atom attached to the β -carbon through the bond that is to be cleaved. If the β -carbon is insufficiently electrophilic or the atom attached to the bond to be cleaved is insufficiently nucleophilic the bond will not be cleaved by a hydrolytic mechanism. Analogs exist which do not mimic the transition state but which are non-hydrolyzable, but transition state isosteres mimic the transition state specifically and are non-hydrolyzable.

Transition state theory indicates that it is the transition state intermediate of the reaction which the enzyme catalyzes for which the enzyme has its highest affinity. It is the transition state structure, not the ground state structure, of the substrate which will have the highest affinity for its given enzyme. The 5 transition state for the hydrolysis of an amide bond is tetrahedral while the ground state structure is planar. A typical transition-state isostere of aspartic protease is -CH(OH)-CH₂-, as was first discovered in pepstatin by Marciniszyn et al. (1976). The transition-state analogue principles have been successfully applied to inhibitor drugs for human immunodeficiency virus protease, an 10 aspartic protease. Many of these are currently in clinical use. Information on the structure, specificity, and types of inhibitors can be found in Tang, Acid Proteases, Structure, Function and Biology, Adv. in Exptl. Med. Biol. vol. 95 (Plenum Press, NY 1977); Kostka, Aspartic Proteinases and their Inhibitors (Walter de Gruyter, Berlin 1985); Dunn, Structure and Functions of the Aspartic 15 Proteinases, Adv. in Exptl. Med. Biol. 306 (Plenum Press, NY 1991); Takahashi, Aspartic Proteases, Structure, Function, Biology, Biomedical Implications, Adv. in Exptl. Med. Biol. 362 (Plenum Press, NY 1995); and James, Aspartic Proteinases, Retroviral and Cellular Enzymes, Adv. in Exptl. Med. Biol. 436 (Plenum Press, NY 1998)).

20 Substrate analog compositions are generally of the general formula X-L₄-P₄-L₃-P₃-L₂-P₂-L₁-P₁-L₀-P₁'-L₁'-P₂'-L₂'-P₃'-L₃'-P₄'L₄'-Y. The substrate analog compositions are analogs of small peptide molecules. Their basic structure is derived from peptide sequences that were determined through structure/function studies. It is understood that positions represented by P_x 25 represent the substrate specificity position relative to the cleavage site which is represented by an -L₀-. The positions of the compositions represented by L_x represent the linking regions between each substrate specificity position, P_x.

In a natural substrate for memapsin 2, a P_x-L_x pair would represent a single amino acid of the peptide which is to be cleaved. In the present general 30 formula, each P_x part of the formula refers to the α -carbon and side chain functional group of each would be amino acid. Thus, the P_x portion of an P_x-L_x

pair for alanine represents HC-CH₃. The general formula representing the P_x portion of the general composition is -R₁CR₃-.

In general R₁ can be either CH₃ (side chain of alanine), CH(CH₃)₂ (side chain of valine), CH₂CH(CH₃)₂ (side chain of leucine), (CH₃)CH(CH₂CH₃) (side chain of isoleucine), CH₂(Indole) (side chain of tryptophan),
5 CH₂(Benzene) (side chain of phenylalanine), CH₂CH₂SCH₃ (side chain of methionine), H (side chain of glycine), CH₂OH (side chain of serine), CHOCH₃ (side chain of threonine), CH₂(Phenol) (side chain of tyrosine), CH₂SH (side chain of cysteine), CH₂CH₂CONH₂ (side chain of glutamine),
10 CH₂CONH₂ (side chain of asparagine), CH₂CH₂CH₂CH₂NH₂ (side chain of lysine), CH₂CH₂CH₂NHC(NH)(NH₂) (side chain of arginine), CH₂(Imidazole) (side chain of histidine), CH₂COOH (side chain of aspartic acid), CH₂CH₂COOH (side chain of glutamic acid), and functional natural and non-natural derivatives or synthetic substitutions of these.

15 It is most preferred that R₃ is a single H. In general, however, R₃ can be alkenyl, alkynal, alkenyloxy, and alkynyloxy groups that allow binding to memapsin 2. Preferably, alkenyl, alkynyl, alkenyloxy and alkynyloxy groups have from 2 to 40 carbons, and more preferably from 2 to 20 carbons, from 2 to 10 carbons, or from 2 to 3 carbons., and functional natural and non-natural derivatives or synthetic substitutions of these.

20

The L_x portion of the P_x-L_x pair represents the atoms linking the P_x regions together. In a natural substrate the L_x represents the β -carbon attached to the amino portion of what would be the next amino acid in the chain. Thus, L_x would be represented by -CO-NH-. The general formula for L_x is represented by R₂. In general R₂ can be CO-HN (amide), CH(OH)(CH₂) (hydroxyethylene), CH(OH)CH(OH) (dihydroxyethylene), CH(OH)CH₂NH (hydroxyethylamine), PO(OH)CH₂ (phosphinate), CH₂NH, (reduced amide). It is understood that more than one L- maybe an isostere as long as the substrate analog functions to inhibit aspartic protease function.

25 30 Ls which are not isosteres may either be an amide bond or mimetic of an amide bond that is non-hydrolyzable.

X and Y represent molecules which are not typically involved in the recognition by the aspartic protease recognition site, but which do not interfere with recognition. It is preferred that these molecules confer resistance to the degradation of the substrate analog. Preferred examples would be amino acids 5 coupled to the substrate analog through a non-hydrolyzable bond. Other preferred compounds would be capping agents. Still other preferred compounds would be compounds which could be used in the purification of the substrate analogs such as biotin.

As used herein, alkyl refers to substituted or unsubstituted straight, 10 branched or cyclic alkyl groups; and alkoxy refers to substituted or unsubstituted straight, branched or cyclic alkoxy. Preferably, alkyl and alkoxy groups have from 1 to 40 carbons, and more preferably from 1 to 20 carbons, from 1 to 10 carbons, or from 1 to 3 carbons.

As used herein, alkenyl refers to substituted or unsubstituted straight 15 chain or branched alkenyl groups; alkynyl refers to substituted or unsubstituted straight chain or branched alkynyl groups; alkenyloxy refers to substituted or unsubstituted straight chain or branched alkenyloxy; and alkynyloxy refers to substituted or unsubstituted straight chain or branched alkynyloxy. Preferably, alkenyl, alkynyl, alkenyloxy and alkynyloxy groups have from 2 to 40 carbons, 20 and more preferably from 2 to 20 carbons, from 2 to 10 carbons, or from 2 to 3 carbons.

As used herein, alkaryl refers to an alkyl group that has an aryl 25 substituent; aralkyl refers to an aryl group that has an alkyl substituent; heterocyclic-alkyl refers to a heterocyclic group with an alkyl substituent; alkyl-heterocyclic refers to an alkyl group that has a heterocyclic substituent.

The substituents for alkyl, alkenyl, alkynyl, alkoxy, alkenyloxy, and 30 alkynyloxy groups can be halogen, cyano, amino, thio, carboxy, ester, ether, thioether, carboxamide, hydroxy, or mercapto. Further, the groups can optionally have one or more methylene groups replaced with a heteroatom, such as O, NH or S.

A number of different substrates were tested and analyzed, and the cleavage rules for Memapsin 2 were determined. The results of the substrates which were analyzed are presented in Table 1 and the rules determined from these results are summarized below.

5 (1) The primary specificity site for a memapsin 2 substrate is subsite position, P_1' . This means that the most important determinant for substrate specificity in memapsin 2 is the amino acid, $S1'$. P_1' must contain a small side chain for memapsin 2 to recognize the substrate. Preferred embodiments are substrate analogs where R_1 of the P_1' position is either H (side chain of glycine), CH_3 (side chain of alanine), CH_2OH (side chain of serine), or CH_2OOH (side chain of aspartic acid). Embodiments that have an R_1 structurally smaller than CH_3 (side chain of alanine) or CH_2OH (side chain of serine) are also preferred.

10 (2) There are no specific sequence requirements at positions P_4 , P_3 , P_2 , P_1 , P_2' , P_3' , and P_4' . Each site can accommodate any other amino acid residue in singularity as long as rule number 3 is met.

15 (3) At least two of the remaining seven positions, P_4 , P_3 , P_2 , P_1 , P_2' , P_3' , and P_4' , must have an R_1 which is made up of a hydrophobic residue. It is preferred that there are at least three hydrophobic residues in the remaining seven positions, P_4 , P_3 , P_2 , P_1 , P_2' , P_3' , and P_4' . Preferred R_1 groups for the positions 20 that contain a hydrophobic group are CH_3 (side chain of alanine), $CH(CH_3)_2$ (side chain of valine), $CH_2CH(CH_3)_2$ (side chain of leucine), $(CH_3)CH(CH_2CH_3)$ (side chain of isoleucine), $CH_2(INDOLE)$ (side chain of tryptophan), $CH_2(Benzene)$ (side chain of phenylalanine), $CH_2CH_2SCH_3$ (side chain of methionine) $CH_2(Phenol)$ (side chain of tyrosine). It is more preferred that the 25 hydrophobic group be a large hydrophobic group. Preferred R_1 s which contain large hydrophobic groups are $CH(CH_3)_2$ (side chain of valine), $CH_2CH(CH_3)_2$ (side chain of leucine), $(CH_3)CH(CH_2CH_3)$ (side chain of isoleucine), $CH_2(Indole)$ (side chain of tryptophan), $CH_2(Benzene)$ (side chain of phenylalanine), $CH_2CH_2SCH_3$ (side chain of methionine) $CH_2(Phenol)$ (side 30 chain of tyrosine). It is most preferred that positions with a hydrophobic R_1 are $CH(CH_3)_2$ (side chain of valine), $CH_2CH(CH_3)_2$ (side chain of leucine),

CH₂(Benzene) (side chain of phenylalanine), CH₂CH₂SCH₃ (side chain of methionine), or CH₂(Phenol) (side chain of tyrosine).

(4) None of the eight positions, P₄, P₃, P₂, P₁, P₁', P₂', P₃', and P₄' may have a proline side chain at its R1 position.

5 (5) Not all subsites must have an P represented in the analog. For example, a substrate analog could have X-P₂-L₁-P₁-L₀-P₁'-L₁'-P₂'-L₂'-P₃'-L₃'-Y or it could have X-L₁-P₁-L₀-P₁'-L₁'-P₂'-L₂'-P₃'-L₃'-P₄'L₄'-Y.

Preferred substrate analogs are analogs having the sequences disclosed in Table 1, with the non-hydrolyzable analog between P1 and P1'.

10 Combinatorial Chemistry to Make Inhibitors

Combinatorial chemistry includes but is not limited to all methods for isolating molecules that are capable of binding either a small molecule or another macromolecule. Proteins, oligonucleotides, and polysaccharides are examples of macromolecules. For example, oligonucleotide molecules with a 15 given function, catalytic or ligand-binding, can be isolated from a complex mixture of random oligonucleotides in what has been referred to as "*in vitro* genetics" (Szostak, *TIBS* 19:89, 1992). One synthesizes a large pool of molecules bearing random and defined sequences and subjects that complex mixture, for example, approximately 10¹⁵ individual sequences in 100 µg of a 20 100 nucleotide RNA, to some selection and enrichment process. Through repeated cycles of affinity chromatography and PCR amplification of the molecules bound to the ligand on the column, Ellington and Szostak (1990) estimated that 1 in 10¹⁰ RNA molecules folded in such a way as to bind a small molecule dyes. DNA molecules with such ligand-binding behavior have been 25 isolated as well (Ellington and Szostak, 1992; Bock et al, 1992).

Techniques aimed at similar goals exist for small organic molecules, proteins and peptides and other molecules known to those of skill in the art. Screening sets of molecules for a desired activity whether based on libraries of small synthetic molecules, oligonucleotides, proteins or peptides is broadly 30 referred to as combinatorial chemistry.

There are a number of methods for isolating proteins either have *de novo* activity or a modified activity. For example, phage display libraries have been used for a number of years. A preferred method for isolating proteins that have a given function is described by Roberts and Szostak (Roberts R.W. and Szostak 5 J.W. Proc. Natl. Acad. Sci. USA, 94(23)12997-302 (1997). Another preferred method for combinatorial methods designed to isolate peptides is described in Cohen et al. (Cohen B.A., et al., Proc. Natl. Acad. Sci. USA 95(24):14272-7 (1998)). This method utilizes a modified two-hybrid technology. Yeast two-hybrid systems are useful for the detection and analysis of protein:protein 10 interactions. The two-hybrid system, initially described in the yeast *Saccharomyces cerevisiae*, is a powerful molecular genetic technique for identifying new regulatory molecules, specific to the protein of interest (Fields and Song, *Nature* 340:245-6 (1989)). Cohen et al., modified this technology so that novel interactions between synthetic or engineered peptide sequences could 15 be identified which bind a molecule of choice. The benefit of this type of technology is that the selection is done in an intracellular environment. The method utilizes a library of peptide molecules that attach to an acidic activation domain. A peptide of choice, for example an extracellular portion of memapsin 2 is attached to a DNA binding domain of a transcriptional activation protein, 20 such as Gal 4. By performing the Two-hybrid technique on this type of system, molecules that bind the extracellular portion of memapsin 2 can be identified.

Screening of Small Molecule Libraries

In addition to these more specialized techniques, methodology well known to those of skill in the art, in combination with various small molecule or 25 combinatorial libraries, can be used to isolate and characterize those molecules which bind to or interact with the desired target, either memapsin 2 or its substrate. The relative binding affinity of these compounds can be compared and optimum inhibitors identified using competitive or non-competitive binding studies which are well known to those of skill in the art. Preferred competitive 30 inhibitors are non-hydrolyzable analogs of memapsin 2. Another will cause

allosteric rearrangements which prevent memapsin 2 from functioning or folding correctly.

Computer assisted Rational Drug Design

Another way to isolate inhibitors is through rational design. This is achieved through structural information and computer modeling. Computer modeling technology allows visualization of the three-dimensional atomic structure of a selected molecule and the rational design of new compounds that will interact with the molecule. The three-dimensional construct typically depends on data from x-ray crystallographic analyses or NMR imaging of the selected molecule. The molecular dynamics require force field data. The computer graphics systems enable prediction of how a new compound will link to the target molecule and allow experimental manipulation of the structures of the compound and target molecule to perfect binding specificity. For example, using NMR spectroscopy, Inouye and coworkers were able to obtain the structural information of N-terminal truncated TSHK (transmembrane sensor histidine kinases) fragments which retain the structure of the individual sub-domains of the catalytic site of a TSHK. On the basis of the NMR study, they were able to identify potential TSHK inhibitors (U.S. Patent No. 6,077,682 to Inouye). Another good example is based on the three-dimensional structure of a calcineurin/FKBP12/FK506 complex determined using high resolution X-ray crystallography to obtain the shape and structure of both the calcineurin active site binding pocket and the auxiliary FKBP12/FK506 binding pocket (U.S. Patent No. 5,978,740 to Armistead). With this information in hand, researchers can have a good understanding of the association of natural ligands or substrates with the binding pockets of their corresponding receptors or enzymes and are thus able to design and make effective inhibitors.

Prediction of molecule-compound interaction when small changes are made in one or both requires molecular mechanics software and computationally intensive computers, usually coupled with user-friendly, menu-driven interfaces between the molecular design program and the user. Examples of molecular modeling systems are the CHARMM and QUANTA programs, Polygen

Corporation, Waltham, MA. CHARMM performs the energy minimization and molecular dynamics functions. QUANTA performs the construction, graphic modeling and analysis of molecular structure. QUANTA allows interactive construction, modification, visualization, and analysis of the behavior of molecules with each other.

5 A number of articles review computer modeling of drugs interactive with specific proteins, such as Rotivinen, et al., 1988 *Acta Pharmaceutica Fennica* 97, 159-166; Ripka, *New Scientist* 54-57 (June 16, 1988); McKinlay and Rossmann, 1989 *Annu. Rev. Pharmacol. Toxicol.* 29, 111-122; Perry and 10 Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis and Dean, 1989 *Proc. R. Soc. Lond.* 236, 125-140 and 141-162; and, with respect to a model enzyme for nucleic acid components, Askew, et al., 1989 *J. Am. Chem. Soc.* 111, 1082-1090. Other computer programs that screen and graphically depict chemicals 15 are available from companies such as BioDesign, Inc., Pasadena, CA., Allelix, Inc, Mississauga, Ontario, Canada, and Hypercube, Inc., Cambridge, Ontario.

20 Although described above with reference to design and generation of compounds which could alter binding, one could also screen libraries of known compounds, including natural products or synthetic chemicals, and biologically active materials, including proteins, for compounds which alter substrate binding or enzymatic activity.

Screening of Libraries

Design of substrate analogs and rational drug design are based on 25 knowledge of the active site and target, and utilize computer software programs that create detailed structures of the enzyme and its substrate, as well as ways they interact, alone or in the presence of inhibitor. These techniques are significantly enhanced with x-ray crystallographic data in hand. Inhibitors can also be obtained by screening libraries of existing compounds for those which inhibit the catalytically active enzyme. In contrast to reports in the literature 30 relating to memapsin 2, the enzyme described herein has activity analogous to the naturally produced enzyme, providing a means for identifying compounds

which inhibit the endogenous activity. These potential inhibitors are typically identified using high throughput assays, in which enzyme, substrate (preferably a chromogenic substrate) and potential inhibitor (usually screened across a range of concentrations) are mixed and the extent of cleavage of substrate determined.

5 Potentially useful inhibitors are those which decrease the amount of cleavage.

II. Methods of diagnosis and treatment

Inhibitors can be used in the diagnosis and treatment and/or prevention of Alzheimer's disease and conditions associated therewith, such as elevated levels of the forty-two amino acid peptide cleavage product, and the

10 accumulation of the peptide in amyloid plaques.

Diagnostic Uses

The substrate analogs can be used as reagents for specifically binding to memapsin 2 or memapsin 2 analogs and for aiding in memapsin 2 isolation and purification or characterization, as described in the examples. The inhibitors

15 and purified recombinant enzyme can be used in screens for those individuals more genetically prone to develop Alzheimer's disease.

Therapeutic Uses

Recombinant human memapsin 2 cleaves a substrate with the sequence LVNM/AEGD (SEQ ID NO:9). This sequence is the *in vivo* processing site

20 sequence of human presenilins. Both presenilin 1 and presenilin 2 are integral membrane proteins. They are processed by protease cleavage, which removes the N terminal sequence from the unprocessed form. Once processed, presenilin forms a two-chain heterodimer (Capell et al., J. Biol. Chem. 273, 3205 (1998);

Thinakaran et al., Neurobiol. Dis. 4, 438 (1998); Yu et al., Neurosci Lett.

25 2,254(3):125-8 (1998)), which is stable relative to the unprocessed presenilins.

Unprocessed presenilins are quickly degraded (Thinakaran et al., J. Biol.

Chem. 272, 28415 (1997); Steiner et al., J. Biol. Chem. 273, 32322 (1998)). It is known that presenilin controls the *in vivo* activity of beta-secretase, which in turn cleaves the amyloid precursor protein (APP) leading to the formation of

30 alpha-beta42. The accumulation of alpha-beta42 in the brain cells is known to be a major cause of Alzheimer's disease (for review, see Selkoe, 1998). The

Memapsin 2 Crystal Structure.

The bilobal structure of memapsin 2 (Figure 7) is characteristic of aspartic proteases (Tang, J., et al., *Nature* 271, 618-621 (1978)) with the conserved folding of the globular core. The substrate binding cleft, where the inhibitor is bound (Figure 7), is located between the two lobes. A pseudo two-fold symmetry between the N- (residues 1-180) and C- (residues 181-385) lobes (Figure 7), which share 61 superimposable atoms with an overall 2.3 Å rms deviation using a 4 Å cutoff. The corresponding numbers for pepsin are 67 atoms and 2.2 Å. Active-site Asp³² and Asp²²⁸ and the surrounding hydrogen-bond network are located in the center of the cleft (Figure 7) and are conserved with the typical active-site conformation (Davies, D. R., *Annu. Rev. Biophys. Chem.* 19, 189 (1990)). The active site carboxyls are, however, not co-planar and the degree of which (50°) exceeds those observed previously.

Compared to pepsin, the conformation of the N-lobe is essentially conserved (Sielecki et al., 1990). The most significant structural differences are the insertions and a C-terminal extension in the C-lobe. Four insertions in helices and loops (Figure 7) are located on the adjacent molecular surface. Insertion F, which contains four acidic residues, is the most negatively charged surface on the molecule. Together, these insertions enlarged significantly the molecular boundary of memapsin 2 as compared to pepsin (Figure 8). These surface structural changes may have function in the association of memapsin 2 with other cell surface components. Insertions B and E are located on the other side of the molecule (Figure 7). The latter contains a beta-strand that paired with part of the C-terminal extension G. A six-residue deletion occurs at position 329 on a loop facing the flap on the opposite side of the active-site cleft, resulting in an apparently more accessible cleft. Most of the C-terminal extension (residues 359-393) is in highly ordered structure. Residues 369-376 form a beta structure with 7 hydrogen bonds to strand 293-299, while residues 378-383 form a helix (Figures 7 and 8). Two disulfide pairs (residues 155/359 and 217/382) unique to memapsin 2 fasten both ends of the extension region to the C-lobe. This C-terminal extension is much longer than those observed

previously and is conformationally different [Cutfield, S. M., et al., Structure 3, 1261 (1995); Abad-Zapatero, C., et al., Protein Sci. 5, 640 (1996); Symersky, J. et al., Biochemistry 36, 12700 (1997); Yang, J., et al., Acta Crystallogr. D 55, 625 (1999)]. The last eight residues (386-393) are not seen in the electron 5 density map; they may form a connecting stem between the globular catalytic domain and the membrane anchoring domain.

Of the 21 putative pro residues only the last six, 43p-48p, are visible in the electron density map. The remainders are likely mobile. Pro-memapsin expressed in mammalian cell culture has an N-terminus position at Glu^{33p}.

10 However, an Arg-Arg sequence present at residues 43p-44p is a frequent signal for pro-protein processing, e.g., in prorenin (Corvol, P. et al., Hypertension 5, 13-9 (1983)). Recombinant memapsin 2 derived from this cleavage is fully active. The mobility of residues 28p-42p suggests that they are not part of the structure of mature memapsin 2.

15 Memapsin 2-OM99-2 Interaction.

The binding of the eight-residue inhibitor OM99-2 in the active-site cleft shares some structural features with other aspartic protease-inhibitor complexes [Davies, D.R., Annu. Rev. Biophys. Chem. 19, 189 (1990); Bailey and Cooper, (1994); Dealwis et al., (1994)]. These include four hydrogen bonds 20 between the two active-site aspartics to the hydroxyl of the transition-state isostere, the covering of the flap (residues 69-75) over the central part of the inhibitor and ten hydrogen bonds to inhibitor backbone (Figure 9). Most of the latter are highly conserved among aspartic proteases [Davies, D. R. Annu. Rev. Biophys. Chem. 19, 189 (1990); Bailey and Cooper, (1994); Dealwis et al., 25 (1994)] except that hydrogen bonds to Gly¹¹ and Tyr¹⁹⁸ are unique to memapsin 2. These observations illustrate that the manner by which memapsin 2 transition-state template for substrate peptide backbone and mechanism of catalysis are similar to other aspartic proteases. These common features are, however, not the decisive factors in the design of specific memapsin 2 30 inhibitors with high selectivity.

The observation important for the design of inhibitor drugs is that the memapsin 2 residues in contact with individual inhibitor side chains (Figure 9) are quite different from those for other aspartic proteases. These side chain contacts are important for the design of tight binding inhibitor with high 5 selectivity. Five N-terminal residues of OM99-2 are in extended conformation and, with the exception of P_1' Ala, all have clearly defined contacts (within 4 Å of an inhibitor side chain) with enzyme residues in the active-site cleft (Figure 9).

The protease S_4 subsite is mostly hydrophilic and open to solvent. The 10 position of inhibitor P_4 Glu side chain is defined by hydrogen bonds to Gly¹¹ and to P_2 Asn (Figure 9) and the nearby sidechains of Arg²³⁵ and Arg³⁰⁷, which explains why the absence of this residue from OM99-2 cause a 10-fold increase in K_i . Likewise, the protease S_2 subsite is relatively hydrophilic and open to solvent. Inhibitor P_2 Asn side chain has hydrogen bonds to P_4 Glu and Arg²³⁵. 15 The relatively small S_2 residues Ser³²⁵ and Ser³²⁷ (Gln and Met respectively in pepsin) may fit a side chain larger than Asn. Memapsin 2 S_1 and S_3 subsites, which consist mostly of hydrophobic residues, have conformations very different from pepsin due to the deletion of pepsin helix hH2 (Dealwis, et al., (1994)). The inhibitor side chains of P_3 Val and P_1 Leu are closely packed 20 against each other and have substantial hydrophobic contacts with the enzyme (Figure 9), especially P_3 interacts with Tyr⁷¹ and Phe¹⁰⁸. In the beta- secretase site of native APP, the P_2 and P_1 residues are Lys and Met respectively. Swedish mutant APP has Asn and Leu in these positions respectively, resulting 25 in a 60-fold increase of k_{cat}/K_m over that for native APP and an early onset of AD described by Mullan, M., et al. [Nat. Genet. 2, 340 (1992)]. The current structure suggests that inhibitor P_2 Lys would place its positively charge in an unfavorable interaction with Arg²³⁵ with a loss of hydrogen bond to Arg²³⁵, while P_1 Met would have less favorable contact with memapsin 2 than does leucine in this site (Figure 10). No close contact with memapsin 2 was seen for 30 P_1' Ala and an aspartic at this position, as in APP, may be accommodated by interacting with Arg²²⁸.

The direction of inhibitor chain turns at P_2' and leads P_3' and P_4' toward the protein surface (Figure 10). As a result, the side-chain position of P_2' Ala deviates from the regular extended conformation. The side chains of P_3' Glu and P_4' Phe are both pointed toward molecular surface with little significant interaction with the protease (Figure 10). The relatively high B-factors (58.2 \AA^2 for Glu and 75.6 \AA^2 for Phe) and less well-defined electron density suggests that these two residues are relatively mobile, in contrast to the defined structure of the S_3' and S_4' subsites in renin-inhibitor (CH-66) complex (Dealwis et al., 1994). The topologically equivalent region of these renin subsites (residues 292 - 297 in pepsin numbering) is deleted in memapsin 2. These observations suggest that the conformation of three C-terminal residues of OM99-2 may be a functional feature of memapsin 2, possibly a way to lead a long protein substrate out of the active-site cleft.

Example 10: Using The Crystal Structure to Design Inhibitors.

Pharmaceutically acceptable inhibitor drugs normally post a size limit under 800 daltons. In the case of memapsin 2 inhibitors, this requirement may even be more stringent due to the need for the drugs to penetrate the blood-brain barrier [Kearney and Aweeka, (1999)]. In the current model, well defined subsite structures spending P_4 to P_2' provide sufficient template areas for rational design of such drugs. The spacial relationships of individual inhibitor side chain with the corresponding subsite of the enzyme as revealed in this crystal structure permits the design of new inhibitor structures in each of these positions. It is also possible to incorporate the unique conformation of subsites P_2' , P_3' and P_4' into the selectivity of memapsin 2 inhibitors. The examples of inhibitor design based on the current crystal structure are given below.

Example A: Since the side chains of P_3 Val and P_1 Leu are packed against each other and there is no enzyme structure between them, cross-linking these side chains would increase the binding strength of inhibitor to memapsin 2. This is because when binding to the enzyme, the cross-linked inhibitors would have less entropy difference between the free and bound forms than their non-cross-linked counterparts [Khan, A.R., et al., Biochemistry, 37, 16839

(1998)]. Possible structures of the cross-linked side chains include those shown in Figure 11.

Example B: The same situation exists between the P4 Glu and P2 Asn. The current crystal structure shows that these side chains are already hydrogen bonded to each other so the cross linking between them would also derive binding benefit as described in the Example A. The cross-linked structures include those shown in Figure 12.

Example C: Based on the current crystal structure, the P1' Ala side chain may be extended to add new hydrophobic, Van der Waals and H-bond interactions.

10 An example of such a design is diagramed in Figure 13.

Example D: Based on the current crystal structure, the polypeptide backbone in the region of P1, P2, and P3, and the side chain of P1-Leu can be bridged into rings by the addition of two atoms (A and B in Figure 14). Also, a methyl group can be added to the beta-carbon of the P1-Leu (Figure 14).

15 Modifications and variations of the methods and materials described herein will be obvious to those skilled in the art and are intended to come within the scope of the appended claims.